Modelling Capacity Scaling of Wireless Social Networks by A Population-Based Social Formation Model

Cheng Wang
Department of Computer Science and Technology
Tongji University
Shanghai, China
3chengwang@gmail.com

Xiang-Yang Li
Department of Computer Science
Illinois Institute of Technology
Chicago, IL, 60616
xli@cs.iit.edu

Changjun Jiang
Department of Computer Science and Technology
Tongji University
Shanghai, China
cjjiang@tongji.edu.cn

ABSTRACT

In this paper, we investigate capacity scaling laws of wireless social networks under the social-based session formation. We model a wireless social network as a three-layered structure, consisting of the *physical layer*, *social layer*, and *session layer*; we propose a cross-layer *distance&density-aware* model, called the *population-based formation model*, under which: 1) for each node v_k, the number of its friends, denoted by q_k, follows a Zipf's distribution with friendship degree clustering exponent γ; 2) q_k *anchor points* are independently chosen according to a probability distribution with density function proportional to $\mathbf{E}_{k,X}^{-\beta}$, where $\mathbf{E}_{k,X}$ is the expected number of nodes (population) within the distance $|v_k - X|$ to v_k, and β is the clustering exponent of friendship formation; 3) finally, q_k nodes respectively nearest to those q_k anchor points are selected as the friends of v_k. We present the density function of general social friendship distribution as the basis for addressing general capacity of wireless social networks. In particular we derive the *social-broadcast* capacity for homogeneous physical layer under both *generalized physical* and *protocol* interference models, taking in account general clustering exponents of both friendship degree and friendship formation in a 2-dimensional parameter space, i.e., $(\gamma, \beta) \in [0, \infty)^2$.

1. INTRODUCTION

Wireless networks are generally the wireless communication implementations for real-life networking applications; then research issues of wireless networks usually come from and aim at the challenges of wireless technology in specific applications, e.g., wireless sensor networks, wireless local area networks, and *wireless social networks*, a wireless implementation of social networks, that is the focus in this work. In social networks, the relationship/edge between users/vertices represents a specific interdependency, such as co-authorship, citationship, or friendship. Based on massive datasets of large-scale real-world online social networks, such as Myspace [1], Twitter [12], Flickr [18], LiveJournal [19], and Facebook [25], extensive studies validate respectively that the two most representative features of *complex networks*, i.e., the *small-world phenomenon* and *scale-free degree distribution*, nearly hold in online social networks. Wireless social networks can be analyzed from a layered perspective, i.e., the social network of users can be regarded as an overlay network over their physical communication network. Therefore, it is necessary for studying wireless social networks to take the property of social networks into account.

As online social networking services are becoming more popular each day and the adoption rates of smart wireless devices like smartphones are increasing aggressively, wireless social network applications will be recognized as the typical instances of large-scale wireless networks. Therefore, it is significant to investigate the fundamental limits of such a large-scale wireless system. As an important metric of fundamental limits, capacity scaling laws of wireless networks, i.e., the scaling of the throughput capacity in the limit when the size of network gets large, have received enormous attentions, [8, 22, 28, 31], since Gupta and Kumar [9] took the lead to study the capacity for homogeneous wireless *random* and *arbitrary* networks. In wireless social networks, source-destination associations depend on their social relationships that had been shown to be inhomogeneous, however, traffic sessions are traditionally assumed to be formed independently in a uniformly random fashion in the literature, which makes the existing results *not* applicable to wireless social networks.This paper aims to introduce the social-based session formation model conforming with wireless social networks, and derive the corresponding capacity scaling laws.

Main Challenges and Our Solutions: We list the following three main challenges in addressing the problem.

Challenge I: Modelling Social Formation Based on Geog-

raphy. On the one hand, since the distance over which data are carried under every session is one of key factors determining the throughput capacity of network, The session formation model should be necessarily relevant to the geography. On the other hand, some classical experiments showed or implied that social relationships among users strongly depend on their geographic location, [7, 10, 17]. Thus, under the assumption that a session is built only among the friends, i.e., the source chooses its destination only from its friends, a geography-based social relationship formation model acts as the precondition for the session formation model. Some theoretical models have been advanced to model the social relationship formation depending on geography in social networks. Kleinberg [11] initiated a *distance-based* social model relating geographical distance and social friendship, in which the probability of befriending a particular user follows a power-law probability distribution versus the distance. This model had been adopted in studying the capacity of wireless social networks, [4, 5]. However, recent experimental studies showed that distance-based models cannot sufficiently capture the characteristics of formation of real-world social networks, [15]. Liben-Nowell et al. [15] stated that the social formation depends on both distance and density, and introduced the *rank-based model*, where the probability of befriending a particular person is inversely proportional to the power of the number of closer people.

To the best of our knowledge, the rank-based model is the most realistic theoretical model for social formation. But it still has shortcoming on analyzing capacity scaling laws in terms of the convenience of analysis or theoretical basis and rigor, because it directly selects nodes instead of points/positions, which leads that the distances of sessions are dependent then causes analysis difficulty in bounding the sum of transport distances of sessions with multiple destinations, [9, 14]. Then, for addressing the capacity scaling laws of wireless networks, it should be the first step to introduce a new distance&density-aware social formation model that is suitable for capacity analysis, while keeping the advantages of rank-based model.

Our Solution for Challenge I: We model a wireless social network as a three-layered structure, consisting of the *physical layer*, *social layer*, and *session layer*, as illustrated in Fig.1. On the basis of rank-based model, we present a cross-layer *distance&density-aware* social model good at the analysis of capacity scaling laws, called the *population-based formation model* $\mathbb{P}(\delta, \gamma, \beta)$, where δ, γ, and β are the clustering exponents of node distribution, friendship degree and friendship formation, respectively. Under $\mathbb{P}(\delta, \gamma, \beta)$: 1) for each node v_k, the number of its friends, denoted by q_k, follows a Zipf's distribution with friendship degree clustering

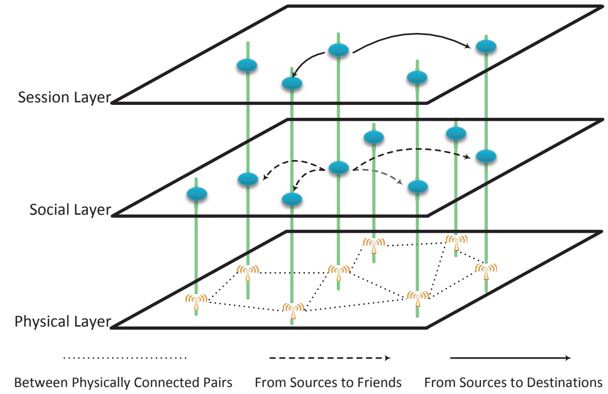


Figure 1: Layered System Model.

exponent γ; 2) q_k *anchor points* are independently chosen according to a probability distribution with density function proportional to $\mathbf{E}_{k,X}^{-\beta}$, where $\mathbf{E}_{k,X}$ is the expected number of nodes (population) within the distance $|v_k - X|$ to v_k, and β is the clustering exponent of friendship formation; 3) finally, select q_k nodes respectively nearest to those q_k anchor points as the friends of v_k. Please refer to Section 2.2.2 for the detailed discussion on the advantages of this model.

Challenge II: Bounding Sum of Distances for All Sessions. The sum of transport distances for all sessions is the prerequisite for bounding the capacity scaling laws, [9, 30]. The long-tailed property of both the destination number (Zipf's Distribution) and destination distribution (Power Law Distribution) causes the transport distances of resulted sessions to be significantly inhomogeneous, which makes it more difficult to bound such a sum.

Our Solution for Challenge II: We present the density function of general social friendships distribution (Theorem 1) and bounds on length of Euclidean spanning trees over nodes chosen according to this density function (Theorem 2). The results can act as the basis for addressing the capacity of wireless social networks.

Challenge III: The complete result under the system model includes the capacity for every point in the 3-dimensional parameter space, i.e., $(\delta, \gamma, \beta) \in [0, \infty)^3$. The complexity of analysis is substantially increased due to the involvement of multiple clustering exponents.

Our Solution for Challenge III: As the first work under this general model, this paper derives the capacity for *social-broadcast* sessions under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, taking in account general clustering exponents of both friendship degree and friendship formation, i.e., $(\gamma, \beta) \in [0, \infty)^2$ (Theorem 3). In addition, we probe the feasibility of studying network capacity under the social model with inhomogeneous physical layer, by extending the basic theorem (Theorem 1) for the general distribution of anchor points.

The rest of the paper is structured as follows. In Section 2,

we introduce the network model. Main results are presented in Section 3. In Section 4, we derive the social-broadcast capacity. We conclude the paper and discuss some topics for future research in Section 5.

2. SYSTEM MODEL

We present a three-layer perspective for the wireless social network with social-based sessions, consisting of the *physical layer*, *social layer* and *session layer*, as in Fig.1.

Throughout the paper, we let $\mathbf{E}[X]$ and $\mathbf{Var}[X]$ denote the mean and variance of a random variable X, respectively.

2.1 Physical Layer Deployment

We introduce a random network model, called the *centre-clustering random model* (CCRM), highlighting the clustering and inhomogeneity property in real-life networks.

2.1.1 Centre-Clustering Random Model (CCRM)

We consider the network composed of a random number of N wireless ad hoc nodes/users distributed over a square region of area $S := n$, where $\mathbf{E}[N] = n$. To avoid border effects, we consider wraparound conditions at the network edges, i.e., the network area is assumed to be the surface of a two-dimensional Torus $\mathcal{O}$. To simplify the description, we assume that the number of nodes is exactly n, and denote the set of nodes by $\mathcal{V} = \{v_k\}_{k=1}^{n}$, without changing our results in order sense, [8, 22, 28].

To emulate the clustering behavior of users distribution in wireless networks, we construct the centre-clustering random model (CCRM) by the following procedure: First, making a center of $\mathcal{O}$ as the centre point, denoted by O. Then, the centre point O generates a point process of nodes whose local intensity at position X is given by

$$\mathbf{d}(X) = n \cdot \kappa(O, X) = n \cdot \frac{g(|X - O|)}{\int_{\mathcal{O}} g(|Y - O|)dY} \quad (1)$$

where $\kappa(O, \cdot)$ is a dispersion density function. As in the literature, we restrict ourselves to the kernel $\kappa(O, \cdot)$ that is invariant under both translation and rotation, i.e., $\kappa(O, X) = \kappa(|X - O|)$ depends only on the Euclidean distance $|X - O|$ of point X from the cluster centre O, [2, 3]. Moreover, we assume that $\kappa(O, \cdot)$ is a summable, non-increasing, bounded and continuous function, and $\int_{\mathcal{O}} \kappa(O, X)dX = 1$. Following a common normalizing method, the kernel can be specified by first defining a non-increasing, bounded and continuous function $g(s)$ and then normalizing it over the area $\mathcal{O}$:

$$\kappa(O, X) = \frac{g(|X - O|)}{\int_{\mathcal{O}} g(|Y - O|)dY}$$

Specifically, we define $g(s) := \min\{1, s^{-\delta}\}$, where $\delta \in [0, \infty)$ is the *clustering exponent of node distribution*.

Note that when $\delta = 0$, the model degenerates into the homogeneous *random extended network*, [8, 26].

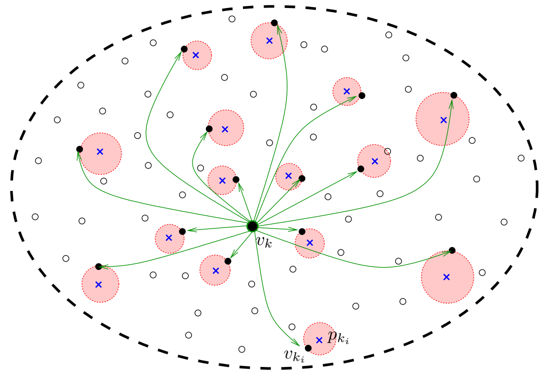


Figure 2: Friendships to Source v_k. Here, point p_{k_i} represents the *anchor point* of node/friend v_{k_i}.

2.1.2 Extending Generality of Physical Layer Model

In this paper, we only study the capacity scaling laws under the CCRM, where there is only one centre point and of extended scaling pattern. We notice that based on the CCRM, one can develop more general physical layer models in terms of *clustering patterns* and *scaling patterns*. To smoothen the mainline of our work, we have moved the details of extended discussion to Section 5.2.

2.2 Social Layer Formation

We introduce a social formation model, called *population-based social model*. We will clarify the advantages of this model later in Section 2.2.2.

2.2.1 Population-based Social Formation Model

Let $\mathcal{D}(u, r)$ denote the disk centered at a point u with radius r in the deployment region $\mathcal{O}$, and let $N(u, r)$ denote the number of nodes contained in $\mathcal{D}(u, r)$.

For a node $v_k \in \mathcal{V}$, construct its friendship set of q_k, $q_k \geq 1$, nodes/friends, say $\mathcal{F}_k$, by the following procedure:

1. Zipf's Degree Distribution of Social Relations: Assume that the number of friends of a particular node $v_k \in \mathcal{V}$, denoted by q_k, follows a Zipf's distribution [16], i.e.,

$$\Pr(q_k = l) = \left(\sum\nolimits_{j=1}^{n-1} j^{-\gamma}\right)^{-1} \cdot l^{-\gamma}. \quad (2)$$

2. Population-Based Formation of Social Relations: Making the position of node v_k as the reference point, choose q_k *points* independently on the torus region $\mathcal{O}$ according to a probability distribution with density function:

$$f_{v_k}(X) = \Phi_k(S, \beta) \cdot (\mathbf{E}[N(v_k, |X - v_k|)] + 1)^{-\beta}, \quad (3)$$

where the random variable $X := (x, y)$ denotes the position of a selected point in the deployment region, $|X - v_k|$ denotes the Euclidean distance between point X and node v_k, $\beta \in [0, \infty)$ represents the clustering exponent of friendship

formation; the coefficient $\Phi_k(S,\beta) > 0$ depends on β and S (the area of deployment region), satisfying that:

$$\Phi_k(S,\beta) \cdot \int_{\mathcal{O}} (\mathbf{E}[N(v_k, |X - v_k|)] + 1)^{-\beta}\, dX = 1. \quad (4)$$

3. Nearest-Principle Position of Friends: Let $\mathcal{A}_k = \{p_{k_i}\}_{i=1}^{q_k}$ denote the set of these q_k points. Let v_{k_i} be the nearest node to p_{k_i}, for $1 \le i \le q_k$ (ties are broken randomly). Denote the set of these q_k nodes by $\mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$. Please see the illustration in Fig.2. We call point p_{k_i} the *anchor point* of v_{k_i}, and define a set $\mathcal{P}_k := \{v_k\} \cup \mathcal{A}_k$.

Throughout this paper, we use $\mathbb{P}(\delta,\gamma,\beta)$ to denote the population-based social model.

2.2.2 Advantages of Population-Based Model

After Kleinberg [11] proposed a *distance-based* social model relating geographical distance and social friendship, Liben-Nowell et al. [15] introduced the *rank-based model*, where the probability of befriending a particular person is inversely proportional to the power of the number of closer people. They validated the practicality of rank-based model by analyzing the data of an online social network, the LiveJournal online community. They [15] pointed out that the weakness of distance-based models lies in that for a particular user, it underestimates the friendship probability of the distant nodes in the low-density region, when the geographical distribution of users is inhomogeneous in common occurrence, as illustrated in Fig.3.

The rank-based model states that the friendship probability depends on both the geographic distance and node density. Following this observation, by modifying the rank-based model, we propose the distance&density-aware population-based social model. We highlight that the population-based model is more convenient and systematic for the issue of capacity scaling laws. Anchor points are usefully introduced, in order to ensure the independence of length of certain Euclidean spanning trees, thus makes it convenient to bound the total length, e.g., the proof of Lemma 9. However, under the rank-based model where the friendships are directly built over nodes without anchor points, the corresponding independence cannot be guaranteed, which usually brings the difficulty on the theoretical rigor. The advantages of the point-based model for the basis and rigor in analysis, compared to the node-based model, had been apparent in [9, 14].

2.3 Session Layer Construction

After the social layer is formed, social sessions can be defined according to the specific applications: For the *social-unicast*/*social-multicast*, the source node delivers message to one/multiple selected friend(s).

For the *social-broadcast*, the source node broadcasts message to all its friends, such as tweets in Twitter and posts in Facebook. Accordingly, we can define other session patterns based on the definitions of corresponding non-social sessions, such as *social-anycast* [13] and *social-manycast* [6].

In this work, we mainly study social-broadcast sessions.

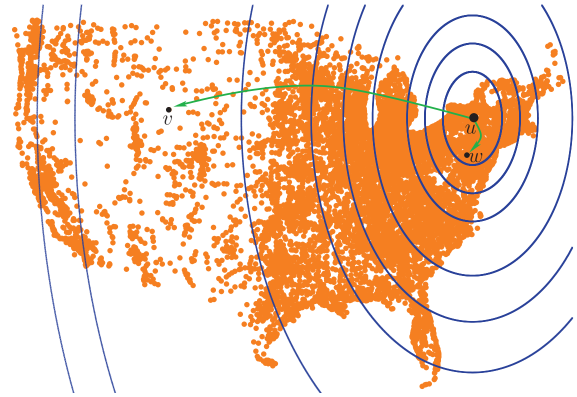


Figure 3: Inhomogeneity of LiveJournal Population [15]. A dot is shown for every distinct United States location home to at least one LiveJournal user (up to Feb. 2004). The population of each successive displayed circle (all centered on Ithaca, NY) increases by 50,000 people. The friendships of $u \to v$ and $u \to w$ are respectively *underestimated* and *overestimated* by the distance-based model.

2.4 Communication Model

We mainly adopt the *generalized physical model* [8, 14] due to its generality and practicality compared to other models like the *protocol model* and *physical model* [9]. Let $\mathcal{L}_t$ denote a *scheduling set* of links in which all links can be scheduled simultaneously in time slot t; let $\alpha > 2$ denote the power attenuation exponent; let B and P denote the bandwidth and transmitting power, respectively.

DEFINITION 1. *Under the generalized physical model, when a scheduling set $\mathcal{L}_t$ is scheduled, the rate of a link $< u, v > \in \mathcal{L}_t$ is achieved at*

$$R_{u,v;t} = B \times \mathbf{1} \cdot \{< u, v > \in \mathcal{L}_t\} \times \log(1 + \mathrm{SINR}_{u,v;t}),$$

where $\mathrm{SINR}_{u,v;t} = \frac{P \cdot \ell(|u-v|)}{N_0 + \sum_{<i,j> \in \mathcal{L}_t / <u,v>} P \cdot \ell(|i-v|)}$, $|u - v|$ *represents the Euclidean distance between nodes u and v; $\ell(\cdot)$ denotes the power attenuation function that is assumed to depend only on the distance between the transmitter and receiver, $\ell(|\cdot|) := |\cdot|^{-\alpha}$ for dense networks, and $\ell(|\cdot|) := \min\{1, |\cdot|^{-\alpha}\}$ for extended networks.*

2.5 Network Capacity for Social Sessions

Denote a session by $\mathcal{S}_k := \{v_k\} \cup \mathcal{D}_k$, where $\mathcal{D}_k \subseteq \mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$ is the set of destinations of v_k. For social-broadcast sessions, it holds that $\mathcal{D}_k = \mathcal{F}_k$.

Let $\Lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ denote a *rate vector* of the data

Table 1: Notations for Exponents

Notation	Definition
$\delta \in [0, \infty)$	clustering exponent of node distribution
$\gamma \in [0, \infty)$	clustering exponent of friendship degree
$\beta \in [0, \infty)$	clustering exponent of friendship formation
$\alpha \in (2, \infty)$	attenuation exponent of signal transmission

rate of all sessions. A rate vector Λ is *feasible* if there is a $T < \infty$ such that in every time interval (with unit seconds) $[(t-1) \cdot T, t \cdot T]$, every source node v_k can send $T \cdot \lambda_k$ bits to all its destinations. For a rate vector, we define the *per-session throughput* as $\Lambda(n) = \min_{v_k \in \mathcal{V}} \lambda_k$.

DEFINITION 2 (ACHIEVABLE THROUGHPUT). *We say a per-session throughput $\Lambda(n)$ is* achievable *for all social sessions if there is a feasible rate vector $\Lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ such that $\Lambda(n) = \min_{v_k \in \mathcal{V}} \lambda_k$.*

DEFINITION 3 (SOCIAL CAPACITY). *The per-session capacity for a class of random networks is of order $\Theta(\Gamma(n))$ if there are constants $0 < \underline{c} < \overline{c} < +\infty$ such that*

$$\lim_{n \to +\infty} \Pr(\Lambda(n) = \underline{c} \cdot \Gamma(n) \text{ is achievable}) = 1,$$
$$\liminf_{n \to +\infty} \Pr(\Lambda(n) = \overline{c} \cdot \Gamma(n) \text{ is achievable}) < 1.$$

3. MAIN RESULTS

To facilitate the reader, we have reported in Table 1 a collection of frequently-used system parameters.

3.1 General Social Points Distribution

In the CCRM, say $\mathcal{N}(n, S; g(\cdot))$, we construct a set of $q+1$ *points*, denoted by $\mathcal{P} = \{X_i\}_{i=0}^q$, by the following procedure: 1. Select arbitrarily a point from $\mathcal{O}$ as the first one in $\mathcal{P}$, denoted by X_0. 2. Making point X_0 as the *reference point* denoted by O', select independently other q points at random according to the probability distribution with density function as described in Eq.(3) (Let $v_k := X_0$).

Let $\mathcal{A} := \{X_i\}_{i=1}^q$. Next, we propose two theorems, and prove them in Appendix B.1.

3.1.1 General Social Distribution Density Function

THEOREM 1. *Making the point X_0 as the reference point O', the distribution of points in $\mathcal{A} = \{X_i\}_{i=1}^q$ follows the probability with density function*

$$f_{X_0}(X) = \frac{\left[\int_{\mathcal{D}(X_0, |X - X_0|)} \mathbf{d}(Y)\, dY + 1\right]^{-\beta}}{\int_{\mathcal{O}} \left[\int_{\mathcal{D}(X_0, |Z - X_0|)} \mathbf{d}(Y)\, dY + 1\right]^{-\beta} dZ} \qquad (5)$$

where $\mathbf{d}(Y) = n \cdot \frac{\min\{1, |Y - O|^{-\delta}\}}{\int_{\mathcal{O}} \min\{1, |Z - O|^{-\delta}\} dZ}$.

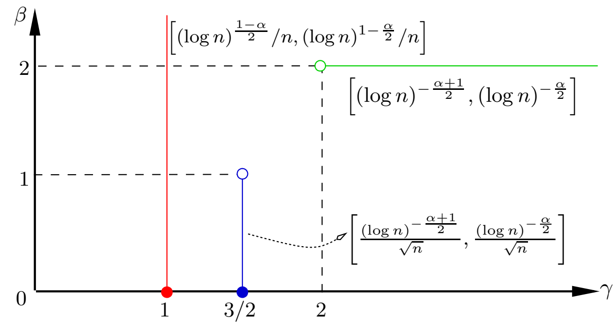


Figure 4: Three Regimes/Lines Where Gaps Exist.

3.1.2 Euclidean Minimum Spanning Tree

THEOREM 2. *Let* $\mathrm{EMST}(\mathcal{P})$ *and* $\mathrm{EMST}(\mathcal{A})$ *denote the Euclidean minimum spanning trees of $\mathcal{P}$ and $\mathcal{A}$, respectively. Then, as $q \to \infty$, with probability 1, it holds that*

$$|\mathrm{EMST}(\mathcal{A})| = \Theta\left(\sqrt{q} \cdot \int_{\mathcal{O}} \sqrt{f_{X_0}(X)} dX\right); \qquad (6)$$

with high probability $1 - o(1/\hat{N})$, it holds that

$$|\mathrm{EMST}(\mathcal{P})| : \left[|\mathrm{EMST}(\mathcal{A})|, |\mathrm{EMST}(\mathcal{A})| + \bar{L}\right], ^{1} \qquad (7)$$

where $f_{X_0}(X)$ is defined in Theorem 1, and

$$\bar{L} = \min\left\{L \,\middle|\, \int_{\mathcal{D}(X_0, L)} f_{X_0}(X) dX = \Omega(\min\{\frac{\log \hat{N}}{q}, 1\})\right\} \qquad (8)$$

with $\hat{N} : (1, n]$ is a given parameter.

Note that the parameter $\hat{N}$ can be defined as the number of nodes with degree of order $\omega(1)$.

3.2 Social-Broadcast Capacity

3.2.1 Capacity under Generalized Physical Model

THEOREM 3. *Under the population-based social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model (GphyM) with $\alpha > 2$, the per-session social-broadcast capacity is of order Λ, where Λ is defined in Table.2.*

From Theorem 3, there are still gaps between upper and lower bounds on social-broadcast capacity under the generalized physical model in three regimes. As illustrated in Fig.4, these three regimes are indeed lines in the 2-dimensional parameter space $(\gamma, \beta) \in [0, 1)^2$. A challenging issue is to close those gaps by presenting possibly new tighter upper and lower bounds by using some new arguments or designing new schemes.

[1] We use the term $f(n) : [\underline{\phi}(n), \overline{\phi}(n)]$ to represent $f(n) = \Omega(\underline{\phi}(n))$ and $f(n) = O(\overline{\phi}(n))$; and use $f(n) : (\underline{\phi}(n), \overline{\phi}(n))$ to represent $f(n) = \omega(\underline{\phi}(n))$ and $f(n) = o(\overline{\phi}(n))$.

Table 2: Social-Broadcast Capacity under GphyM

γ	Social Capacity under GphyM – Λ:
$\gamma > 2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}), & \beta > 2; \\ \left[(\log n)^{-\frac{\alpha+1}{2}}, (\log n)^{-\frac{\alpha}{2}}\right], & \beta = 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha+3}{2}}), & \beta \ge 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma}), & \beta \ge 2\gamma - 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}/\sqrt{n}), & \beta \ge 1; \\ \left[\frac{(\log n)^{-\frac{\alpha+1}{2}}}{\sqrt{n}}, \frac{(\log n)^{-\frac{\alpha}{2}}}{\sqrt{n}}\right], & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma})$
$\gamma = 1$	$\boxed{(\log n)^{\frac{1-\alpha}{2}}/n, (\log n)^{1-\frac{\alpha}{2}}/n}$
$0 \le \gamma < 1$	$\Theta((\log n)^{-\frac{\alpha}{2}}/n)$

3.2.2 Capacity under Protocol Model

We concentrate on deriving the capacity under the generalized physical model, while for completeness, we also include the results on capacity under the well-known protocol model (ProM, [9]). Based on our results on EMSTs and ESTs of social-broadcast sessions (Theorem 4 and Theorem 7), using the analytical methods for capacity under the protocol model in [9,30], we can obtain that under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the ProM, the per-session social-broadcast capacity is of order Λ^{Pro} as defined in Table.3.

Table 3: Social-Broadcast Capacity under ProM: Λ^{Pro}

γ	Social Capacity under ProM – Λ^{Pro}
$\gamma > 2$	$\begin{cases} \Theta(1/\log n), & \beta > 2; \\ \Theta(1/(\log n)^{\frac{3}{2}}), & \beta = 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Theta(1/(\log n)^3), & \beta \ge 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta(n^{\gamma-2}/\log n), & \beta \ge 2\gamma - 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2\gamma - 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta(1/(\log n \cdot \sqrt{n})), & \beta \ge 1; \\ \Theta(1/(\log n \cdot \sqrt{n \log n})), & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta(1/(\log n \cdot n^{2-\gamma}))$
$0 \le \gamma \le 1$	$\Theta(1/n)$

4. SOCIAL-BROADCAST CAPACITY

4.1 System Settings

In this work, the first one under this model, we specifically reduce the complexity from three dimensions $(\delta, \gamma, \beta) \in [0, \infty)^3$ to two dimensions $(\gamma, \beta) \in [0, \infty)^2$ by letting $\delta = 0$. In this case of extremely weak clustering behavior, the physical layer degenerates into the homogeneous random network model, [8,9,27], where $\mathbf{d}(Y) \equiv \Theta(1)$.

Considering the degree distribution, by Eq.(2), we get that

$$\Pr(q_k = l) = \begin{cases} \Theta(l^{-\gamma}), & \gamma > 1; \\ \Theta(\frac{1}{\log n} \cdot l^{-1}), & \gamma = 1; \\ \Theta(n^{\gamma-1} \cdot l^{-\gamma}), & 0 \le \gamma < 1. \end{cases} \quad (9)$$

4.1.1 Distribution of Anchor Points

For each session $\mathcal{S}_k$ initiated by the source v_k, we can get the distribution of anchor points directly using Theorem 1,

LEMMA 1. *When the clustering exponent $\delta = 0$, for a session $\mathcal{S}_k$ under the population-based social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, the anchor points of the friends of source v_k follows the distribution of density function:*

$$f_{v_k}(X) = \begin{cases} \Theta((|X - v_k|^2 + 1)^{-\beta}), & \beta > 1; \\ \Theta(\frac{1}{\log n} \cdot (|X - v_k|^2 + 1)^{-1}), & \beta = 1; \\ \Theta(n^{\beta-1} \cdot (|X - v_k|^2 + 1)^{-\beta}), & 0 \le \beta < 1. \end{cases}$$

By using Lemma 1, we can get the following result.

LEMMA 2. *For a social-broadcast session $\mathcal{S}_k$ under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, it holds that:*

$$\mathbf{E}[|X - v_k|] = \begin{cases} \Theta(1), & \beta > 3/2; \\ \Theta(\log n), & \beta = 3/2; \\ \Theta(n^{\frac{3}{2}-\beta}), & 1 < \beta < 3/2; \\ \Theta(\sqrt{n}/\log n), & \beta = 1; \\ \Theta(\sqrt{n}), & 0 \le \beta < 1. \end{cases} \quad (10)$$

4.1.2 Social-Broadcast Sessions

Under the population-based social model, we denote a social-broadcast session by a set $\mathcal{S}_k = \mathcal{P}_k := \{v_k\} \cup \mathcal{F}_k$, where v_k is the source node and each element in $\mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$, say v_{k_i}, is the nearest *node* to the corresponding *anchor point* p_{k_i} in $\mathcal{A}_k = \{p_{k_i}\}_{i=1}^{q_k}$. Please see the illustration in Fig.2. Recall that $\mathcal{P}_k = \{v_k\} \cup \mathcal{A}_k$, we get the following Lemma 3 for spanning trees over $\mathcal{S}_k$.

LEMMA 3. *For a social-broadcast session $\mathcal{S}_k$ with $q_k = \omega(1)$ under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, with probability 1, it holds that $|\operatorname{EMST}(\mathcal{A}_k)| = \Theta(L_{\mathcal{P}}(\beta, q_k))$, and then $|\operatorname{EMST}(\mathcal{P}_k)| = \Omega(L_{\mathcal{P}}(\beta, q_k))$, where*

$$L_{\mathcal{P}}(\beta, q_k) = \begin{cases} \Theta(\sqrt{q_k}), & \beta > 2; \\ \Theta(\sqrt{q_k} \cdot \log n), & \beta = 2; \\ \Theta(\sqrt{q_k} \cdot n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(\sqrt{q_k} \cdot \sqrt{\frac{n}{\log n}}), & \beta = 1; \\ \Theta(\sqrt{q_k} \cdot \sqrt{n}), & 0 \le \beta < 1. \end{cases} \tag{11}$$

PROOF. From Theorem 2, it follows that with probability 1, $|\operatorname{EMST}(\mathcal{A}_k)| = \Theta(L_{\mathcal{P}}(\beta, q_k))$ for $q_k = \omega(1)$, where $L_{\mathcal{P}}(\beta, q_k)$ is defined in Eq.(11). Combining with the fact that $|\operatorname{EMST}(\mathcal{P}_k)| \ge |\operatorname{EMST}(\mathcal{A}_k)|$, we get the lemma. □

4.2 Upper Bounds on Social Capacity

4.2.1 Techniques for Upper Bounds

We adopt a technique based on *lattice view* [26] to compute the upper bounds.

DEFINITION 4 (LATTICE VIEW). *Partition a square deployment region $\mathcal{O}(S) = [0, \sqrt{S}]^2$ into $\lceil \sqrt{S}/c \rceil^2$ cells of side length c:$[\sqrt{S/n}, \sqrt{S})$, we call the produced lattice graph* lattice view*, and denote it by $\mathbb{V}(\sqrt{S}, c)$.*

Based on a given lattice view $\mathbb{V}(\sqrt{S}, c)$, we can get the following lemma for arbitrary routing trees for a session $\mathcal{S}_k$. By Lemma 2 of [26], we have

LEMMA 4. *Given a social-broadcast session $\mathcal{S}_k$, let $\mathcal{T}_k$ be a routing tree for $\mathcal{S}_k$, and let $N(\mathcal{T}_k, \sqrt{S}, c)$ denote the number of cells used by $\mathcal{T}_k$ in $\mathbb{V}(\sqrt{S}, c)$, then when $q_k = O(S/c^2)$, it holds that $N(\mathcal{T}_k, \sqrt{S}, c) = \Omega(\frac{1}{c} \cdot |\operatorname{EMST}(\mathcal{S}_k)|)$.*

In $\mathbb{V}(\sqrt{S}, c)$, a cell is called an *island* if it contains $\Theta(\frac{nc^2}{S})$ nodes and all its eight neighbor cells are empty.

LEMMA 5 ([26]). *There exists w.h.p. an island in the lattice view $\mathbb{V}(\sqrt{S}, c)$, if $c \le \frac{1}{2} \cdot \sqrt{\frac{(1-\epsilon) \cdot S \cdot \log n}{2n}}$, where $\epsilon \in (0, 1)$ is constant.*

4.2.2 Lower Bounds on $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{S}_k)|$

Since the sum of length of Euclidean minimum spanning trees for all n sessions, i.e., $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{S}_k)|$, plays a key role in the analysis of network capacity, we first give the lower bounds on $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{S}_k)|$. Note that the bounds depend on those on $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{P}_k)|$, which are provided in Lemma 9 in Appendix B.2.

THEOREM 4. *For all social-broadcast sessions $\mathcal{S}_k$, $k = 1, 2, \cdots, n$, under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, with high probability, $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{S}_k)| = \Omega(H(\gamma, \beta))$, where $H(\gamma, \beta)$ is described in Table.4.*

PROOF. Please see the proof in Appendix B.2. □

Table 4: $H(\gamma, \beta)$ in Bounding $\sum \operatorname{EMST}$ and $\sum \operatorname{EST}$

γ	$H(\gamma, \beta)$
$\gamma > 2$	$\begin{cases} \Theta(n), & \beta > 2; \\ \Theta(n \cdot \log n), & \beta = 2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Theta(n \cdot \log n), & \beta \ge 2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta(n^{3-\gamma}), & \beta \ge 2\gamma - 2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta(n^{3/2}), & \beta > 1; \\ \Theta(n^{3/2} \cdot \sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2} \cdot \log n), & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta(n^{3-\gamma})$
$\gamma = 1$	$\Theta(n^2/\log n)$
$0 \le \gamma < 1$	$\Theta(n^2)$

4.2.3 Upper Bound from Lattice View $\mathbb{V}(\sqrt{n}, \sqrt{2})$

We adopt the lattice view $\mathbb{V}(\sqrt{n}, \sqrt{2})$ to derive an upper bound of the social-broadcast capacity. For completeness, we include a useful result from [26]:

LEMMA 6 ([26]). *The throughput capacity of any cell in lattice view $\mathbb{V}(\sqrt{n}, \sqrt{2})$ is at most of order $O(1)$.*

Combining Lemma 6 and Theorem 4, we can obtain

THEOREM 5. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model with $\alpha > 2$, the per-session social-broadcast capacity is of order $O(n/H(\gamma, \beta))$, where $H(\gamma, \beta)$ is defined in Table.4.*

PROOF. For each routing tree $\mathcal{T}_k$, denote the number of cells in $\mathbb{V}(\sqrt{n}, \sqrt{2})$ used by $N(\mathcal{T}_k, \sqrt{n}, \sqrt{2})$. From Lemma 4, $\sum_{k=1}^{n} N(\mathcal{T}_k, \sqrt{n}, \sqrt{2}) = \Omega\left(\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{S}_k)|\right)$. Combining with Theorem 4, we get that

$$\sum\nolimits_{k=1}^{n_s} N(\mathcal{T}_k, \sqrt{n}, \sqrt{2}) = \Omega(H(\gamma, \beta)),$$

where $H(\gamma, \beta)$ is presented in Table.4. By pigeonhole principle, there is at least one cell that will be used by at least $\Omega(H(\gamma, \beta)/n)$ sessions. By Lemma 6, the total throughput capacity of any cell in $\mathbb{V}(\sqrt{n}, \sqrt{2})$ is of order $O(1)$. Thus, under any strategy, due to the congestion in some cells, the multicast throughput is at most of order $O(n/H(\gamma, \beta))$. □

4.2.4 Upper Bound from Lattice View $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$

From Lemma 5, for $\epsilon = \frac{1}{9}$, there is an island in the lattice view $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$. Thus, we get the following theorem.

THEOREM 6. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model with $\alpha > 2$, the per-session social-broadcast capacity is of order*

$$\Lambda = \begin{cases} O((\log n)^{-\frac{\alpha}{2}}), & \gamma > 2; \\ O((\log n)^{-\frac{\alpha}{2}-1}), & \gamma = 2; \\ O(n^{\gamma-2} \cdot (\log n)^{-\frac{\alpha}{2}}), & 1 < \gamma < 2; \\ O((\log n)^{1-\frac{\alpha}{2}}/n), & \gamma = 1; \\ O((\log n)^{-\frac{\alpha}{2}}/n), & 0 \le \gamma < 1. \end{cases}$$

PROOF. Denote an island in $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$ by $\mathcal{I}$. For a link, say $u \to v$, where the receiver v is located in $\mathcal{I}$, its length is $|uv| = \Omega(\sqrt{\log n})$, then the capacity of this link is not larger than $B\log_2(1 + \frac{P\cdot|uv|^{-\alpha}}{N_0}) = O((\log n)^{-\frac{\alpha}{2}})$. Consider the initial transmission load of $\mathcal{I}$, by the *tails of Binomial distribution* and *union bounds*, we have that with uniformly high probability, the initial transmission load of each cell in $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$, denoted by $\mathrm{Load_I}$, is of order

$$\mathrm{Load_I} = \begin{cases} \Omega(\log n), & \gamma > 2; \\ \Omega((\log n)^2), & \gamma = 2; \\ \Omega(\log n \cdot n^{2-\gamma}), & 1 < \gamma < 2; \\ \Omega(n), & \gamma = 1; \\ \Omega(n \cdot \log n), & 0 \le \gamma < 1. \end{cases}$$

In addition, there are at most $\Theta(\log n)$ simultaneous links terminating (or initiating) in $\mathcal{I}$ since it contains $\Theta(\log n)$ nodes inside. By the *pigeonhole principle*, there exists a link whose load is of $\Omega(\mathrm{Load_I}/\log n)$. Then, the lemma follows from $R_{u,v} = O((\log n)^{-\frac{\alpha}{2}})$. □

4.2.5 Combination of Upper Bounds

Combining Theorem 5 and Theorem 6, we obtain the upper bounds in Theorem 3 by performing some simple algebraic manipulations.

4.3 Lower Bounds on Social Capacity

In this section, we present the constructive lower bounds for the social-broadcast capacity by devising two social-broadcast strategies having their own merits respectively. Both strategies depend on the Euclidean spanning trees (ESTs) of sessions and percolation-based routing backbones [8].

4.3.1 Construction of Euclidean Spanning Trees

For each session $\mathcal{S}_k = \{v_k\} \cup \mathcal{D}_k$, where $\mathcal{D}_k = \mathcal{F}_k$, we build an EST, denoted as $\mathrm{EST}(\mathcal{S}_k)$, by the following method (*Anchors-EMST-Based Greedy* (AEBG) Algorithm):

1. Construct an EMST based on $\mathcal{A}_k$, denoted as $\mathrm{EMST}(\mathcal{A}_k)$, using some classic greedy algorithms like Prim algorithm.

2. Connect the pairs v_{k_i} and v_{k_j} if and only if the link $p_{k_i}p_{k_j} \in \mathrm{EMST}(\mathcal{A}_k)$. Then, one can obtain an EST of $\mathcal{F}_k$.

3. Connect the source v_k to its nearest node in $\mathcal{F}_k$ to get the final EST of $\mathcal{S}_k$, i.e., $\mathrm{EST}(\mathcal{S}_k)$.

Next, we give the upper bounds on $\sum_{k=1}^{n} |\mathrm{EST}(\mathcal{S}_k)|$.

THEOREM 7. *For all social-broadcast sessions $\mathcal{S}_k$, $k = 1, 2, \cdots, n$, under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, using the aforementioned AEBG algorithm, with high probability, it holds that $\sum_{k=1}^{n} |\mathrm{EST}(\mathcal{S}_k)| = O(H(\gamma, \beta))$, where $H(\gamma, \beta)$ is described in Table.4.*

PROOF. Please see the proof in Appendix B.3. □

Combining Theorem 4 and Theorem 7, we get that

PROPOSITION 1. *For all social-broadcast sessions $\mathcal{S}_k$, under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, it holds that with high probability, $\sum_{k=1}^{n} |\mathrm{EMST}(\mathcal{S}_k)| = \Theta(H(\gamma, \beta))$, where $H(\gamma, \beta)$ is described in Table.4.*

4.3.2 Social-Broadcast Schemes

Since the physical layer with $\delta = 0$ is a *random extended network*, the percolation-based routing backbone [8] still applies to the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$. Since a social-broadcast session is similar to the multicast session for the source, We can draw on the design of multicast strategies. To the best of our knowledge, the tightest lower bound on multicast capacity for random extended networks was derived in [27] by designing two multicast schemes based on two types of hierarchical backbones systems in a well-integrated manner: The former's hierarchical routing backbone consists of *highways* [8] and *parallel arterial roads* [27], denoted by $\mathbb{M}_{\mathrm{p\&h}}$; The latter is only based on *parallel arterial roads*, denoted by $\mathbb{M}_{\mathrm{p}}$. Due to limited space, we cannot include the complete descriptions of these schemes. Please refer to [27].

Next, we give a theorem to present the achievable social-broadcast throughputs under $\mathbb{M}_{\mathrm{p\&h}}$ and $\mathbb{M}_{\mathrm{p}}$.

THEOREM 8. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and generalized physical model with $\alpha > 2$, using $\mathrm{EST}(\mathcal{S}_k)$ derived by AEBG algorithm as the input of schemes $\mathbb{M}_{\mathrm{p\&h}}$ or $\mathbb{M}_{\mathrm{p}}$, then it holds that with high probability:*

$\triangleright$ *Under scheme $\mathbb{M}_{\mathrm{p\&h}}$, the achievable throughput, denoted by $\underline{\Lambda}^{\mathbb{M}_{\mathrm{p\&h}}}$, is described in Table.5.*

$\triangleright$ *Under scheme $\mathbb{M}_{\mathrm{p\&h}}$, the achievable throughput, denoted by $\underline{\Lambda}^{\mathbb{M}_{\mathrm{p}}}$, is described in Table.6.*

Combining the throughputs under schemes $\mathbb{M}_{\mathrm{p\&h}}$ and $\mathbb{M}_{\mathrm{p}}$ in Theorem 8, we get the lower bounds in Theorem 3.

4.3.3 Proof of Theorem 8

Define an event $\mathcal{E}_k(\mathbb{M}, v)$ as: Session $\mathcal{S}_k$ is routed through v under the scheme $\mathbb{M}$ in the corresponding phase.

Table 5: Achievable Throughput under $\mathbb{M}_{\text{p\&h}}$: $\underline{\Lambda}^{\mathbb{M}_{\text{p\&h}}}$

γ	$\underline{\Lambda}^{\mathbb{M}_{\text{p\&h}}}$
$\gamma > 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha+1}{2}}), & \beta \geq 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha+3}{2}}), & \beta \geq 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Omega(\frac{(\log n)^{-\frac{\alpha+1}{2}}}{n^{2-\gamma}}), & \beta \geq 2\gamma - 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$1 < \gamma \leq 3/2$	$\Omega((\log n)^{-\frac{\alpha+1}{2}}/n^{2-\gamma})$
$\gamma = 1$	$\Omega((\log n)^{\frac{1-\alpha}{2}}/n)$
$0 \leq \gamma < 1$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n)$

Table 6: Achievable Throughput under $\mathbb{M}_{\text{p}}$: $\underline{\Lambda}^{\mathbb{M}_{\text{p}}}$

γ	$\underline{\Lambda}^{\mathbb{M}_{\text{p}}}$
$\gamma > 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}), & \beta > 2; \\ \Omega((\log n)^{-\frac{\alpha+1}{2}}), & \beta = 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}-2}), & \beta \geq 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma}), & \beta \geq 2\gamma - 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}/\sqrt{n}), & \beta \geq 1; \\ \Omega((\log n)^{-\frac{\alpha+1}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$1 \leq \gamma < 3/2$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma})$
$0 \leq \gamma < 1$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n)$

$\triangleright$ *Under Scheme* $\mathbb{M}_{\text{p\&h}}$: For a highway-station [27], say v^{H}, we can get that $\Pr(\mathcal{E}_k(\mathbb{M}_{\text{p\&h}}, v)) = O(\frac{1}{n} \cdot (|\operatorname{EST}(\mathcal{S}_k)| + q_k \cdot \log n))$. By laws of larger numbers, the load of highway-station v^{H} in the *highway phase*, denoted by $\operatorname{Load}(v^{\text{H}})$, holds that $\operatorname{Load}(v^{\text{H}}) = O(\frac{1}{n} \cdot (\sum_{k=1}^{n} |\operatorname{EST}(\mathcal{S}_k)| + \min\{\log n \cdot \sum_{k=1}^{n} q_k, n^2\}))$. Combining the fact that v^{H} can sustain the rate of order $\Theta(1)$ in highway phase, we get the throughput during highway phase as follows:

$$\underline{\Lambda}^{\text{H}} = \Omega(n/(H(\gamma, \beta) + \min\{\log n \cdot Q(\gamma), n^2\})), \quad (12)$$

where $H(\gamma, \beta)$ and $Q(\gamma)$ are described in Table.4 and Eq.(18), respectively. Similarly, we get the throughput in *AR phase*,

$$\underline{\Lambda}^{\text{AR}} = \Omega(\lambda^{\text{AR}} \cdot n / \min\{\sqrt{\log n} \cdot Q(\gamma), n^2\}), \quad (13)$$

where $\lambda^{\text{AR}} = \Omega((\log n)^{-\frac{\alpha}{2}})$ is the rate of parallel AR-station in AR phase [27].

According to the principle of bottleneck, combining Eqs.(12) and (13), we obtain the achievable throughput $\underline{\Lambda}^{\mathbb{M}_{\text{p\&h}}}$.

$\triangleright$ *Under Scheme* $\mathbb{M}_{\text{p}}$: For a parallel AR station, say v^{pa}, it can sustain the rate of order $\Omega((\log n)^{-\frac{\alpha}{2}})$. By a similar analysis method, we get the achievable throughput $\underline{\Lambda}^{\mathbb{M}_{\text{p}}}$ by

$$\underline{\Lambda}^{\mathbb{M}_{\text{p}}} = \Omega((\log n)^{\frac{1-\alpha}{2}} \cdot n/(H(\gamma, \beta) + \sqrt{\log n} \cdot Q(\gamma))). \quad (14)$$

5. CONCLUSION AND FUTURE WORK

In this paper, we mainly address capacity scaling laws of wireless social networks under social-based session formation. A three-layered model is proposed for abstracting wireless social networks. As one of main contributions, a cross-layer and distance&density-aware social model is proposed, which captures the formation characteristics of real-world social networks better, and specializes in the analysis of capacity scaling laws. We derive the social-broadcast capacity, taking in account the general clustering exponents of both friendship degree and friendship formation. Moreover, we present the density function of general social friendships distribution that will be the basis for investigating the capacity of general wireless social networks.

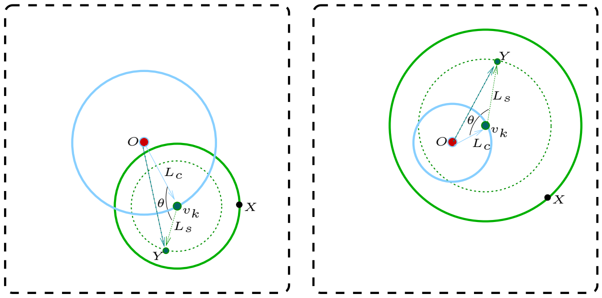
5.1 General Physical Clustering Case

The advantages of population-based model cannot be sufficiently highlighted for the case that $\delta = 0$, indeed. It would be a significant future work to investigate the relationships between the general clustering exponent and capacity in 3-dimensional parameter space, i.e., $(\delta, \gamma, \beta) \in [0, \infty)^3$.

Here, we probe the feasibility of studying network capacity under the social model with inhomogeneous physical layer, by extending the proposed basic theorem (Theorem 1) for the general distribution of anchor points. The key factor is to determine $\mathbf{d}(Y)$. Furthermore, since $\mathbf{d}(Y) = \frac{n \cdot \min\{1, |Y-O|^{-\delta}\}}{\int_{\mathcal{O}} \min\{1, |Z-O|^{-\delta}\} dZ}$, then it should be the first key step to deal with $|Y - O|$ based on the "known" *distance deviating from the source* $L_s = |Y - v_k|$ and *distance deviating from the centre* $L_c = |v_k - O|$. As illustrated in Fig.5, it follows that $|Y - O| = \sqrt{L_c^2 + L_s^2 - 2L_c \cdot L_s \cdot \cos\theta}$, which can provide a basis for the further study of general model.

5.2 Extending to General Physical Layer

The proposed CCRM can be developed to more general physical layer models in terms of *clustering patterns* and *scaling patterns*.



(a) $|X - v_k| \le |v_k - O|$ (b) $|X - v_k| > |v_k - O|$

Figure 5: Illustrations of positions of the centre point O, the reference point v_k, the targeted point X, and any interior point Y in $\mathcal{D}(v_k, |X - v_k|)$.

$\triangleright$ *Multiple Clustering Centres*– In the CCRM, there is only one centre. Although the results under this model also apply to the model with physical layer model having a finite number of centres due to the characteristic of capacity scaling issues, more realistic and general clustering behaviors in real-life networks cannot be fully embodied by this simple model. To model the clustering behavior of node distribution with $\omega(1)$ centres in real-life applications of wireless networks, the shotnoise Cox process (SNCP, [2,20]) can be introduced, where M centre points generate their respective point processes, and the conditional local intensity at X is determined by the superposition of the individual processes, i.e., $\mathbf{d}(X) = \sum_j \rho_j \cdot \kappa(c_j, X)$, Then, the node process forms an inhomogeneous Poisson point process.

$\triangleright$ *General Scaling Model*– In the research of network capacity scaling laws, there are two typical models in terms of scaling patterns of network: *dense scaling model* and *extended scaling model* [8, 21, 27]. They have different engineering implications related to the classical notions of *interference-limitedness* and *coverage-limitedness* [21]: the former is only interference-limited; while, the latter is both interference-limited and coverage-limited. In the CCRM, the area of deployment region is $S := n$, which implies that it is of *extended scaling pattern*. Then, it is necessary to extend the CCRM to one with general scaling pattern by setting a general area of deployment region.

6. REFERENCES

[1] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proc. ACM WWW 2007*.

[2] G. Alfano, M. Garetto, and E. Leonardi. Capacity scaling of wireless networks with inhomogeneous node density: Upper bounds. *IEEE Journal on Selected Areas in Comm.*, 27(7):1147–1157, 2009.

[3] G. Alfano, M. Garetto, E. Leonardi, and V. Martina. Capacity scaling of wireless networks with inhomogeneous node density: Lower bounds. *IEEE/ACM Trans. on Networking*, 18(5):1624–1636, 2010.

[4] B. Azimdoost and H. Sadjadpour. Capacity of scale free wireless networks. In *Proc. IEEE GlobeCom 2012*.

[5] B. Azimdoost, H. Sadjadpour, and J. Garcia-Luna-Aceves. Capacity of composite networks: Combining social and wireless ad hoc networks. In *Proc. IEEE WCNC 2011*.

[6] C. Carter, S. Yi, P. Ratanchandani, and R. Kravets. Manycast: exploring the space between anycast and multicast in ad hoc networks. In *Proc. ACM MobiCom 2003*.

[7] P. Dodds, R. Muhamad, and D. Watts. An experimental study of search in global social networks. *science*, 301(5634):827–829, 2003.

[8] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran. Closing the gap in the capacity of wireless networks via percolation theory. *IEEE Trans. on Information Theory*, 53(3):1009–1018, 2007.

[9] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Trans. on Information Theory*, 46(2):388–404, 2000.

[10] P. Killworth and H. Bernard. The reversal small-world experiment. *Social networks*, 1(2):159–192, 1979.

[11] J. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.

[12] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proc. ACM WWW 2010*.

[13] V. Lenders, M. May, and B. Plattner. Density-based anycast: a robust routing strategy for wireless ad hoc networks. *IEEE/ACM Trans. on Networking*, 16(4):852–863, 2008.

[14] S. Li, Y. Liu, and X.-Y. Li. Capacity of large scale wireless networks under Gaussian channel model. In *Proc. ACM MobiCom 2008*.

[15] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *PNAS of the United States of America*, 102(33):11623–11628, 2005.

[16] C. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

[17] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[18] A. Mislove, H. Koppula, K. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *Proc. ACM Workshop on Online social networks 2008*.

[19] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proc. ACM SIGCOMM Conference on Internet Measurement 2007*.

[20] J. Møller. Shot noise cox processes. *Advances in Applied Probability*, 35(3):614–640, 2003.

[21] A. ÖzgÜr, O. LÉvÊque, and D. Tse. Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks. *IEEE Trans. on Information Theory*, 53(10):3549–3572, 2007.

[22] Q. Peng, X. Wang, and H. Tang. Heterogeneity increases multicast capacity in clustered network. In *Proc. IEEE INFOCOM 2011*.

[23] M. Penrose. A strong law for the longest edge of the minimal spanning tree. *The Annals of Probability*, 27(1):246–260, 1999.

[24] J. Steele. Growth rates of euclidean minimal spanning trees with power weighted edges. *The Annals of Probability*, pages 1767–1787, 1988.

[25] B. Viswanath, A. Mislove, M. Cha, and K. Gummadi. On the evolution of user interaction in facebook. In *Proc. ACM Wrkshop on Online Social Networks 2009*.

[26] C. Wang, C. Jiang, X.-Y. Li, S. Tang, Y. He, X. Mao, and Y. Liu. Scaling laws of multicast capacity for power-constrained wireless networks under gaussian channel model. *IEEE Trans. Computers*, 61(5):713–725, 2012.

[27] C. Wang, C. Jiang, X.-Y. Li, S. Tang, and P. Yang. General capacity scaling of wireless networks. In *Proc. IEEE INFOCOM 2011*.

[28] X. Wang, L. Fu, and C. Hu. Multicast performance with hierarchical cooperation. *Networking, IEEE/ACM Transactions on*, 20(3):917–930, 2012.

[29] D. Williams. *Probability with martingales*. Cambridge university press, 1991.

[30] L. Xiang-Yang., S. Tang, and F. Ophir. Multicast capacity for large scale wireless ad hoc networks. In *Proc. ACM MobiCom 2007*.

[31] L. Ying, S. Yang, and R. Srikant. Optimal delay–throughput tradeoffs

in mobile ad hoc networks. *Information Theory, IEEE Transactions on*, 54(9):4119–4143, 2008.

APPENDIX

A. USEFUL KNOWN RESULTS

LEMMA 7 (MINIMAL SPANNING TREE, [24]). *Let X_i, $1 \leq i < \infty$, denote independent random variables with values in $\mathbb{R}^d$, $d \geq 2$, and let M_n denote the cost of a minimal spanning tree of a complete graph with vertex set $\{X_i\}_{i=1}^n$, where the cost of an edge (X_i, X_j) is given by $\Psi((|X_i - X_j|))$. Here, $|X_i - X_j|$ denotes the Euclidean distance between X_i and X_j and Ψ is a monotone function. For bounded random variables and $0 < \sigma < d$, it holds that as $n \to \infty$, with probability 1, one has $M_n \sim c_1(\sigma, d) \cdot n^{\frac{d-\sigma}{d}} \cdot \int_{\mathbb{R}^d} f(X)^{\frac{d-\sigma}{d}} dX$,provided $\Psi(x) \sim x^\sigma$, where $f(X)$ is the density of the absolutely continuous part of the distribution of the $\{X_i\}$.*

LEMMA 8 (KOLMOGOROV'S STRONG LLN, [29]). *Let $\{X_n\}$ be an i.i.d. sequence of random variables having finite mean: for $\forall n$, $\mathbf{E}[X_n] < \infty$. Then, a strong law of large numbers (LLN) applies to the sample mean: $\bar{X}_n \xrightarrow{a.s.} \mathbf{E}[X_n]$, where $\xrightarrow{a.s.}$ denotes* almost sure convergence.

B. PROOFS OF SOME THEOREMS

B.1 Proof of Theorem 1 and Theorem 2

First of all, the cost of an edge (X_i, X_j) is given by $\Psi(|X_i - X_j|) = |X_i - X_j|$, that is, the exponent σ in Lemma 7 equals 1. In addition, $\Psi(x)$ is a monotonically increasing function. Let L denote the distance between the centre O and reference point. Then, under the centre-clustering random model $\mathcal{N}(n, S; g(\cdot))$, by Eqs.(3) and (4), the density function is specified into Eq.(5). Then, by Lemma 7, we get that

$$|\operatorname{EMST}(\mathcal{A})| = \Theta\left(\sqrt{q} \cdot \int_{\mathcal{O}} \sqrt{f_{X_0}(X)} dX\right).$$

It is straightforward that $|\operatorname{EMST}(\mathcal{P})| = \Omega(|\operatorname{EMST}(\mathcal{A})|)$. On the other hand, let $\underline{L}$ denote the smallest distance from the points in $\mathcal{A}$ to point X_0. Then,

$$\left(1 - \int_{\mathcal{D}(X_0, \underline{L})} f_{X_0}(X) dX\right)^q = o(1).$$

That is, $\int_{\mathcal{D}(X_0, \underline{L})} f_{X_0}(X) dX = \omega(1/q)$. Thus, $\underline{L} \leq \bar{L}$, where $\bar{L}$ is defined in Eq.(8), which completes the proof. Note that we deliberately relax the upper bound of $\underline{L}$ as in Eq.(8) in order to ensure Eq.(7) to hold with uniformly high probability for $\Theta(n)$ Euclidean spanning trees, [23].

B.2 Proof of Theorem 4

First, we give a basic lemma for the final proof.

LEMMA 9. *For all social-broadcast sessions $\mathcal{S}_k$ ($k = 1, 2, \cdots, n$) under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, with high probability, the lower bounds on $\sum_{k=1}^n |\operatorname{EMST}(\mathcal{P}_k)|$ hold as described in Table.7.*

PROOF. Let N_l denote the number of sessions with l destinations. First of all, to simplify the proof, we let

$$N_l = n \cdot \Pr(q_k = l) = n \cdot \left(\sum_{j=1}^{n-1} j^{-\gamma}\right)^{-1} \cdot l^{-\gamma},$$

which has no impact on the analysis in order sense according to laws of larger numbers. Based on all $\mathcal{S}_k$, define two sets $\mathcal{K}^1 := \{k | q_k = \Theta(1)\}$ and $\mathcal{K}^\infty := \{k | q_k = \omega(1)\}$. Then,

$$\sum_{k=1}^n |\operatorname{EMST}(\mathcal{P}_k)| = \underline{\Sigma}^1 + \underline{\Sigma}^\infty, \tag{15}$$

where $\underline{\Sigma}^1 = \sum_{k \in \mathcal{K}^1} |\operatorname{EMST}(\mathcal{P}_k)|$, $\underline{\Sigma}^\infty = \sum_{k \in \mathcal{K}^\infty} |\operatorname{EMST}(\mathcal{P}_k)|$.

First, we consider $\underline{\Sigma}^1$. Sine for $q_k = \Theta(1)$, it holds that $|\operatorname{EMST}(\mathcal{P}_k)| = \Theta(|X - v_k|)$, then we have $\underline{\Sigma}^1 = \sum_{k \in \mathcal{K}^1} |X - v_k|$. For $k \in \mathcal{K}^1$, define a sequence of random variables $\xi_k^1 := |X - v_k| / \sqrt{n}$ having finite mean: $\mathbf{E}[\xi_k^1] = \mathbf{E}[|X - v_k|] / \sqrt{n}$, where $\mathbf{E}[|X - v_k|]$ is presented in Lemma 2. Then, $\underline{\Sigma}^1 = \Theta(\sqrt{n} \cdot \sum_{k \in \mathcal{K}^1} \xi_k^1)$. Therefore, by Lemma 8, with probability 1, $\sum_{k \in \mathcal{K}^1} \xi_k^1 = \Theta(|\mathcal{K}^1| \cdot \mathbf{E}[|X - v_k| / \sqrt{n}])$, where $|\mathcal{K}^1|$ denotes the cardinality of $\mathcal{K}^1$. Thus, we get that with probability 1,

$$\underline{\Sigma}^1 = \Theta(|\mathcal{K}^1| \cdot \mathbf{E}[|X - v_k|]). \tag{16}$$

Next, we consider $\underline{\Sigma}^\infty$. For $k \in \mathcal{K}^\infty$, all random variables $|\operatorname{EMST}(\mathcal{P}_k)|$ are *independent*; moreover, from Lemma 3, with probability 1, $|\operatorname{EMST}(\mathcal{P}_k)| = \Omega(L_{\mathcal{P}}(\beta, q_k))$, where $L_{\mathcal{P}}(\beta, q_k)$ is defined in Eq.(11). Thus, with probability 1,

$$\underline{\Sigma}^\infty \geq \sum_{l:(1,n]} n \cdot (\sum_{j=1}^{n-1} j^{-\gamma})^{-1} \cdot l^{-\gamma} \cdot L_{\mathcal{P}}(\beta, l). \tag{17}$$

Finally, combining Eqs.(15), (16) and (17), we complete the proof. □

Then, we begin to prove Theorem 4. Since $\sum_{k=1}^n q_k = \Theta(\sum_{l=1}^{n-1} n \cdot \Pr(q_k = l) \cdot l)$, we get $\sum_{k=1}^n q_k = Q(\gamma)$, where

$$Q(\gamma) = \begin{cases} \Theta(n), & \gamma > 2; \\ \Theta(n \log n), & \gamma = 2; \\ \Theta(n^{3-\gamma}), & 1 < \gamma < 2; \\ \Theta(n^2 / \log n), & \gamma = 1; \\ \Theta(n^2), & 0 \leq \gamma < 1. \end{cases} \tag{18}$$

Moreover, for all $v_k \in \mathcal{V}$, $\mathbf{E}[|v_{k_i} - p_{k_i}|] = \Theta(\int_0^{\sqrt{n}} x \cdot e^{-\pi \cdot x^2} dx)$, that is, $\mathbf{E}[|v_{k_i} - p_{k_i}|] = \Theta(1)$. Thus, according to Lemma 8, with high probability,

$$\sum_{k=1}^n \sum_{i=1}^{q_k} |v_{k_i} - p_{k_i}| = \Theta(\sum_{k=1}^n q_k). \tag{19}$$

Finally, combining with Lemma 9, we get the theorem.

Table 7: **Lower Bounds on $\sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{P}_k)|$**

$\beta \backslash \gamma$	$\gamma > 3/2$	$\gamma = 3/2$	$1 < \gamma < 3/2$	$\gamma = 1$	$0 \le \gamma < 1$
$\beta > 2$	$\Omega(n)$	$\Omega(n \log n)$	$\Omega(n^{\frac{5}{2}-\gamma})$	$\Omega(n^{3/2}/\log n)$	$\Omega(n^{3/2})$
$\beta = 2$	$\Omega(n \cdot \log n)$	$\Omega(n \cdot (\log n)^2)$	$\Omega(\log n \cdot n^{\frac{5}{2}-\gamma})$	$\Omega(n^{3/2})$	$\Omega(n^{3/2} \cdot \log n)$
$1 < \beta < 2$	$\Omega(n^{2-\frac{\beta}{2}})$	$\Omega(n^{2-\frac{\beta}{2}} \cdot \log n)$	$\Omega(n^{\frac{7}{2}-\gamma-\frac{\beta}{2}})$	$\Omega(n^{(5-\beta)/2}/\log n)$	$\Omega(n^{(5-\beta)/2})$
$\beta = 1$	$\Omega(n^{3/2}/\sqrt{\log n})$	$\Omega(n^{3/2} \cdot \sqrt{\log n})$	$\Omega(n^{3-\gamma}/\sqrt{\log n})$	$\Omega(n^2/(\log n)^{3/2})$	$\Omega(n^2/\sqrt{\log n})$
$0 \le \beta < 1$	$\Omega(n^{3/2})$	$\Omega(n^{3/2} \cdot \log n)$	$\Omega(n^{3-\gamma})$	$\Omega(n^2/\log n)$	$\Omega(n^2)$

B.3 Proof of Theorem 7

By the proposed construction of EST, we have that

$$\sum\nolimits_{k=1}^{n} \operatorname{EST}(\mathcal{S}_k) = O(\Sigma^{\mathcal{A}} + \Sigma^{vp} + \Sigma^{r}) \qquad (20)$$

where $\Sigma^{\mathcal{A}} = \sum_{k=1}^{n} |\operatorname{EMST}(\mathcal{A}_k)|$, $\Sigma^{vp} = \sum_{k=1}^{n} \sum_{i=1}^{q_k} |v_{k_i} - p_{k_i}|$ and $\Sigma^{r} = \sum_{k=1}^{n} \min_i\{|v_k - v_{k_i}|\}$. Let $\Sigma^{\mathcal{A}} = \overline{\Sigma}^1 + \overline{\Sigma}^\infty$, where $\overline{\Sigma}^1 = \sum\limits_{k \in \mathcal{K}^1} |\operatorname{EMST}(\mathcal{A}_k)|$, $\overline{\Sigma}^\infty = \sum\limits_{k \in \mathcal{K}^\infty} |\operatorname{EMST}(\mathcal{A}_k)|$.

By similar procedures to Eqs.(16) and (17), we obtain that

$$\begin{aligned}
\Sigma^{r} &= O(n \cdot \mathbf{E}[|X - v_k|]), && (21)\\
\overline{\Sigma}^1 &= O(|\mathcal{K}^1| \cdot \mathbf{E}[|X - v_k|]), && (22)\\
\overline{\Sigma}^\infty &= \sum\nolimits_{l:(1,n]} \Big(\sum\nolimits_{j=1}^{n-1} \frac{1}{j^\gamma}\Big)^{-1} \cdot \frac{n}{l^\gamma} \cdot L_{\mathcal{P}}(\beta, l), && (23)\\
\Sigma^{vp} &= \Theta\Big(\sum\nolimits_{k=1}^{n} q_k\Big). && (24)
\end{aligned}$$

Combining Eq.(20) and Eqs.(21-24), we get the theorem.